Journal of Computational and Applied Mathematics 300 (2016) 207–216

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

A new efficient conjugate gradient method for unconstrained optimization

Masoud Fatemi *

Department of Mathematics, K. N. Toosi University of Technology, Tehran, Iran
Scientific Computations in Optimization and System Engineering (SCOPE), K. N. Toosi University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 2 July 2015
Received in revised form 14 November 2015

Keywords:
Conjugate gradient method
Dai–Liao family
Unconstrained optimization
Line search

ABSTRACT

We propose a nonlinear conjugate gradient method for unconstrained optimization based on solving a new optimization problem. Our optimization problem combines the good features of the linear conjugate gradient method using some penalty parameters. We show that the new method is a subclass of Dai–Liao family, the fact that enables us to analyze the family, closely. As a consequence, we obtain an optimal bound for Dai–Liao parameter. The global convergence of the new method is investigated under mild assumptions. Numerical results show that the new method is efficient and robust, and outperforms CG-DESCENT.

1. Introduction

Here, we consider the following unconstrained optimization problem,

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1}$$

where f is a smooth function. Conjugate gradient algorithms are a class of efficient methods for solving (1), specially, when n is large [1–9]. This class was originally invented by Hestenes and Stiefel [1] for solving a symmetric and positive definite linear system of equations, and then was extended by many authors to handle general optimization problems [10,11].

In a conjugate gradient algorithm, a sequence of iterates, x_{k+1}, are generated by the following scheme:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where the search direction d_k is computed by

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \quad d_0 = -g_0. \tag{3}$$

The step length $\alpha_k > 0$ usually satisfies the Wolfe conditions,

$$f(x_{k+1}) - f(x_k) \le c_1 \alpha_k g_k^T d_k, \tag{4}$$

$$g_{k+1}^T d_k \ge c_2 g_k^T d_k, \tag{5}$$

where, $0 < c_1 < c_2 < 1$ are some arbitrary constants and $g_k := \nabla f(x_k)$.

The linear conjugate gradient method uses (2) and (3) with the exact line search to solve a strongly convex quadratic function. The method has some remarkable properties:

* Correspondence to: Department of Mathematics, K. N. Toosi University of Technology, Tehran, Iran.
E-mail address: smfatemi@kntu.ac.ir.

http://dx.doi.org/10.1016/j.cam.2015.12.035
0377-0427/

(i) The sufficient descent condition, namely, there exists a scalar $c > 0$ such that

$$g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2. \tag{6}$$

(ii) The conjugacy condition, namely,

$$d_{k+1}^T y_k = 0. \tag{7}$$

(iii) The orthogonality property:

$$g_{k+1}^T d_i = 0, \tag{8}$$

for $i = 0 \dots k$.

It is known that a linear conjugate gradient algorithm always terminates in finite iteration, and it is actually a remarkable property. Unfortunately, this property cannot be guaranteed for the nonlinear conjugate gradient algorithms.

The good features of the linear conjugate gradient method persuade the authors to follow this idea in nonlinear optimization. We refer the interested readers to the nice surveys of Hager and Zhang [10] and Dai [11] about nonlinear conjugate gradient methods. Hager and Zhang [7] have recently introduced an efficient nonlinear conjugate gradient method. Their method is a subclass of Dai–Liao family corresponding to the choice

$$\beta_k^{DL} = \frac{g_{k+1}^T y_k - \tau g_{k+1}^T s_k}{y_k^T d_k}, \tag{9}$$

with $\tau = \lambda_k \frac{\|y_k\|^2}{s_k^T y_k}$. They proved the global convergence of a truncated version of the method similar to PRP$^+$ of Gilbert and Nocedal [12] under mild assumptions. Numerical results showed that the method outperforms many existing conjugate gradient methods. Nowadays, it is known as a most efficient conjugate gradient method. An implementation of the method called CG-DESCENT is now available from the Hager's homepage.

In order to design an efficient nonlinear conjugate gradient method, we combine (i)–(iii), and introduce the following optimization problem:

$$\min_{\beta_k}\left[g_{k+1}^T d_{k+1} + M\left((g_{k+2}^T s_k)^2 + (d_{k+1}^T y_k)^2\right)\right], \tag{10}$$

where M is a penalty parameter. The first term in (10) contains the information about (i), whereas, the second one contains the information about (ii) and (iii). A large value of M clearly increases the chance of satisfying (ii) and (iii), and a small one increases the effect of the sufficient descent property (i).

More recently in [13], we used the same idea, and proposed an optimal parameter for Dai–Liao family of conjugate gradient methods. In this work, we pay attention to M and introduce an efficient penalty parameter, having some useful properties. A new expression for β_k is then obtained by solving (10). We show that the resulting method is a subclass of Dai–Liao family. This observation enables us to closely analyze Dai–Liao family. As a consequence, we show that the optimal Dai–Liao parameter should be somewhere in interval $(0, \frac{1}{2})$. We also investigate the global convergence of the method under suitable assumptions. Finally, we show that the new method is efficient, and outperforms CG-DESCENT.

The paper is organized as follows: In Section 2, we present some motivations and background of the method. Introducing penalty parameter M is the subject of Section 3. The global convergence of the new method is investigated in Section 4, and numerical results are reported in Section 5. Finally, conclusions and discussions are made in the last section.

2. Motivations and backgrounds

In this section, we introduce a new family of conjugate gradient methods by solving (10).

To solve (10), we should replace g_{k+2} by some of its appropriate estimation, because it is not available in the current iteration.

Here, we consider the quadratic approximation of the objective function,

$$\Phi(d) = f_{k+1} + g_{k+1}^T d + \frac{1}{2} d^T B_{k+1} d,$$

and, take $\nabla\Phi(\alpha_{k+1} d_{k+1})$ as an estimation of g_{k+2}. It is easy to see that

$$\nabla\Phi(\alpha_{k+1} d_{k+1}) = \alpha_{k+1} B_{k+1} d_{k+1} + g_{k+1}. \tag{11}$$

Now, we modify (11) and set

$$g_{k+2} := t B_{k+1} d_{k+1} + g_{k+1}, \tag{12}$$

where, $t > 0$ is a suitable approximation of α_{k+1}. Substituting (3) and (12) in (10), we obtain

$$\beta_k = \frac{1}{X}\left[-g_{k+1}^T d_k + 2Mt^2(s_k^T B_{k+1} g_{k+1})(s_k^T B_{k+1} d_k) - 2Mt(s_k^T g_{k+1})(s_k^T B_{k+1} d_k) + 2M(y_k^T g_{k+1})(y_k^T d_k)\right], \tag{13}$$